SCHOOL OF AVIATION MEDICINE
U.S. NAVAL AIR STATION
PENSACOLA FLORIDA

The Reliability of The Verhoeff Test of Depth Perception

Research Project X-717(Av-374-w) Report No. One

3 May 1946

NAVAL SCHOOL OF AVIATION MEDICINE
U. S. NAVAL AIR TRAINING BASES
PENSACOLA, FLORIDA

RESEARCH REPORT

Submitted - 3 May 1946

PROJECT NO: X-717(Av-374-w)

REPORT NO. ONE

TITLE: THE RELIABILITY OF THE VERHOEFF TEST OF DEPTH PERCEPTION.

REPORT Richard Trumbull
Lt. Richard Trumbull, H(S), USNR

APPROVED: Louis Iverson.
Captain Louis Iverson, (MC), USN
Medical Officer-in-Charge

Summary:

The reliability of the Verhoeff test of depth perception was determined by a test-retest study of one hundred male subjects. This study introduces a variation in presentation and scoring for this device. Four scoring methods were studied for relative reliability and discrimination between levels of depth perception. Statistical analysis of these methods includes: (1) deriving test-retest coefficients of reliability and (2) obtaining measures of dispersion for the methods studied.

Conclusions:

1. Coefficients of test-retest reliability of the Verhoeff Stereopter, as obtained by means of four different scoring methods, were .79, .81, .82 and .82.

2. The four scoring methods studied indicate that this device may be used for testing with various levels of discrimination depending upon requirements established without significantly lowering reliability.

3. In ease of administration, control of variables, and scoring, it is one of the best of present devices for testing stereopsis.

Introduction:

Binocular acuity or stereoscopic vision is determined chiefly by three factors:(1)

1. The possession of two foveas which are corresponding points.
2. The semidecussation of the optic nerve fibers.
3. A certain amount of disparity of the two retinal images.

The last named factor is the only one which is subject to the control necessary in experimentation. For this reason, the testing of stereopsis has been limited to experimental design which presented objects whose retinal images had the necessary disparity. Thus, tests for this function have been based upon finding the greatest disparity imperceptible at a given distance or the least disparity perceptible at that distance. Assuming the latter to be of greater importance, Verhoeff has constructed a test which emphasizes perception of least relative depth.(2)

Another factor in the mechanism of stereopsis is the nature of the fusion movements of the eyes. Bielschowsky (3) presents the findings of his study of these movements as follows:

"There are three pairs of fusion movements; convergence and divergence, positive and negative vertical divergence, and conclination and disclination. The only fusion movement which -- at least to a certain extent -- can be performed voluntarily, is that of convergence, because the latter is also a link in the mechanism of near vision, which is governed by the will. In people with normal binocular vision, all the other fusion movements take place only if the identical images are shifted from corresponding to disparate areas of the two retinas."

Betts (4) has divided fusion further into three levels or degrees, the third of which constitutes the visual perception of solidity and depth.

Verhoeff long has maintained that this perception, dependent upon the conversion of true binocular parallax into depth meaning, takes place only below the level of consciousness. His Stereopter is constructed to reduce acuity of depth perception testing to this pure form. He contends that any test based upon a comparison of two objects of equal size or upon perception of rate of change in binocular parallax, both of which are conscious functions, is open to question. In the Stereopter, he not only has omitted such questionable cues but deliberately has introduced misleading monocular

criteria in order to "make binocular parallax the only correct evidence of relative depth and to cause perception of false depth when this correct evidence was not perceived."(2) A size difference between test objects is the misleading monocular cue employed. Appreciating the fact that judgments as to relative depth of objects in common experience rarely are made with two objects of equal size, it is surprising that this variable has not received wider application. Other controls include:

1. Uniform illumination eliminating brightness contrast cues.
2. Development of an experimental design which included eliminating any cues in the face of the testing screen or positions of the test objects which lend themselves to interpretation at a more perceptible level than do the test objects themselves.

Procedure:

Experimental Design - A test-retest study of stereopsis was conducted with the Verhoeff Stereopter to determine the reliability of that device. Retests were scheduled after a minimum interval of 48 hours; the group average interval was 78 hours. A check-off form was used for indicating responses. Various methods were studied to determine which method gave the highest reliability and served to discriminate adequately between levels of depth perception. Three positions, 2 meters, 1 meter and ½ meter, were established as testing distances.

Subjects - There were 100 male subjects used in this study. An effort was made to obtain a representative group of acuity values, as indicated by scores on the Grow chart.(5) The population was taken from:

1. Volunteers from the officer and enlisted personnel attached to the Main Dispensary at Pensacola.
2. Students from a class of Hospital Corpsmen under training as Aviation Medicine Technicians.
3. Enlisted personnel appearing for refractions at the Eye Clinic prior to administration of homatropine. These subjects were retested after the effects of homatropine had worn off.

The age range represented by this group was from 18.5 to 45 years with a median of 20.25 years, mode of 19 years and mean age of 22.75 years.

Equipment and Testing Method - The only equipment used in this

study was the Verhoeff Stereopter (2), a device for measuring acuity of stereopsis in the absence of any viewing instrument. It consists of a small box attached to a rectangular black target screen, approximately 9 by 17.5 cm. in size. A target window, 1 by 5.4 cm., is centered across the screen. All edges of the window are sharp; side edges are beveled toward the front while the top and bottom edges are beveled toward the back.

"Immediately behind this window, held so it can slide only vertically, is a small screen (sliding screen) 11 cm. high, 6.9 cm. wide and exactly 2.5 mm. thick. The sliding screen contains four rectangular windows, each 16 by 50 mm. in size. These are centered on the vertical midline with their long axes horizontal and are separated from each other by distances of 5 mm. Crossing each window vertically are three thin black strips, 3mm., 2.5mm. and 2mm. in width, respectively. Of the strips, some are affixed to the back and others to the front of the sliding screen. There is, therefore, a depth of 2.5mm. between the strips at the front and those at the back. By moving the sliding screen, one can expose any of the four sets of strips in the target window, and by turning the device upside down, one can reverse the positions of the lateral strips. -In each set the 3mm. strip is centered exactly on the midline, while at one side the 2.5mm. strip is centered 10.75mm. and at the other side the 2mm. strip is centered 10.50mm. from the midline. In set 1, the middle strip is at the front, the 2.5mm. strip at the back on the left and the 2mm. strip at the front on the right. In set 2, the middle strip is at the back, the 2.5mm. strip at the back on the right and the 2mm. strip at the front on the left. In set 3, the middle strip is at the back, the 2.5mm. strip at the front on the left and the 2mm. strip at the back on the right. In set 4, the middle strip is at the back, the 2.5mm. strip at the front on the right and the 2mm. strip at the front on the left.

"Behind the target window and about 3mm. behind the sliding screen is a translucent diffusing screen of ample size (5.8 by 2 cm.). This is indirectly attached to the target screen and is therefore stationary."(2)

The source of illumination is a 2 volt flashlight bulb drawing on two 1.5 volt flashlight batteries. The bulb and batteries are housed in the small container which clips onto the back of the large target screen and is removed easily for necessary replacements. The back section of the screen not covered by this box contains the button used in positioning the test strips. The positions are identified by letters; MF, LN, LN, and LF for one presentation and RF, RN, RN, and MF for the other when the device is inverted. This method allows presentation of eight trials employing the strips

described above. The first of the two letters indicates the test stick, whether the Middle, Left or Right stick, which is displaced. The last letter indicates the displacement as being "Nearer" or "Farther" than the other two sticks.

Testing Method - Preliminary to the testing, the Stereopter was held ½ m. from the subject at eye level. Settings were used indiscriminately during these demonstrations. Instructions were: "Here you see three sticks. You are asked to tell which one seems to be nearest to you and which one seems to be farthest from you - unless they all appear to be at the same distance." If the subject identified the setting properly, the examiner stated: "You will note that the width of the sticks does not enter into your judgment. Remember this during the test proper." If the subject did not identify the setting properly, the examiner allowed further trials after stating: "These three sticks are not of the same width se you must not use apparent difference in width as a means for judging their distance from you."

The Stereopter was held at a distance of 2 m. at eye level. The examiner held it in his right hand and about 1 foot to his right side. In this position, the target was in front of a black background formed by a dark shade. Special care was exercised by the examiner to prevent rotation of the target on any axis. This method of holding the target places the examiner in such a position that he can ascertain that the target is at eye-level for the subject and is not rotated. Four settings were presented in random order, the target being lowered to change setting and to record the response after each judgment. The target then was inverted for presentation of the remaining settings following the same procedure. If but one or two errors were made at 2 m., the settings in question were shown again after presentation of other settings. The target never was held in position following a judgment for further analysis by the subject or to obtain a correct response. The same procedure was followed whether the subject was right or wrong in his judgment in order to eliminate any pressure or establish any false sense of security. The test was terminated if all eight settings were identified correctly. If any errors were made, the subject moved forward to the 1 m. position. The same procedure was followed at this distance with failers moved forward to the ½ m. position. No subject failed to pass at this distance.

Scoring Methods - Four scoring methods were studied for relative reliability and ability to discriminate between levels of stereopsis. These were:

1. A simple pass-fail scoring, using a perfect score at 2 meters as passing.

2. A composite score obtained by crediting each judgment correct at 2 meters with 4, each one correct at 1 meter with 2, and each one correct at ½ meter with 1. When all judgments at one level were correct, credit was given for lower levels without testing.

3. A score based upon order of difficulty. A distribution of errors at the 2 meter level resulted in the following tabulation for the eight settings.

Set 1	Set 2	Set 3	Set 4
LF	LN	LN	MF
32	53	41	39
Set 8	**Set 7**	**Set 6**	**Set 5**
RF	RN	RN	MF
25	41	30	38

During the testing program, it was noted that two sets, 2 and 7, were proving to be more difficult than the other six. These sets involved the same stick in the same depth relationship to the other two sticks. This was the 2mm. stick which was nearer on the left with the testing screen held in one position and nearer on the right when the screen was inverted. Whether this was due to reluctance to accept binocular parallax cues in place of relative size as criterion is not certain although such a situation was anticipated in the construction of the device. Arbitrarily, these two judgments were assigned a value of 2 when correct while all others remained at a value of 1. This rather crude system of equating was tried merely to determine its possibilities if carried to the extreme of giving values equal to actual percentage of error for each setting as determined on a larger sample.

4. A score derived by crediting each correct judgment as 1. Thus, a perfect score at 2 meters received 24 credits. Credit was given for lower levels without testing whenever the subject received a perfect score at a higher level.

Results:

The first of the scoring methods employed, that of a pass-fail dichotomy, gave a tetra-choric coefficient of reliability of .79. The statistical data for the other three methods follow:

	Scoring Method		
	2	3	4
r	.81	.82	.82
M retest minus M test	3.32	1.46	.99
Total possible score	56	30	24
σ_r	.03	.03	.03

It is evident that there is no significant difference in reliability between scoring methods and that each method is significantly reliable within itself.

Discussion:

Generally speaking, tests of stereopsis either depend upon a judgment as to relative position of objects in a fixed field or termination of movement of an object or objects in relation to a fixed criterion or to each other. This test is one of the former. Two other tests of the same type are included in the Bausch and Lomb Ortho-Rater and the Keystone Telebinocular. Coefficients of reliability reported for these tests include .83 for the Ortho-Rater from a study (6) employing a population similar to that of this study and .80 for the Telebinocular using a group of college students (4). The Howard-Dolman test is the most widely used of the second type. Others have been developed for selection purposes in service programs but they are quite similar in respect to fundamentals. One study of the Howard-Dolman test which involves manipulation of two objects into positions of apparent equality gave the following coefficients of reliability for various orders of presentation: (6)

Method of Measurement	Reliability
Average setting, "from behind"	.69
Average setting, "in front"	.78
Median setting, "from behind"	.69
Median setting, "in front"	.75
Variability score, "from behind"	.72
Variability score, "in front"	.75

The technique for this study of the Verhoeff Stereopter differed from that of the originator of the device in three ways:

1. The device was held at the examiner's side for more accurate placement relative to eye-level of the subject and prevention of rotation about any axis.
2. There were three predetermined positions for presentation.
3. None of the scoring systems employed here correspond to that used by Verhoeff.

As presented, this test of stereopsis demonstrates a reliability equal to or better than that of most other tests of this function. Of the four scoring systems employed, none has a significantly higher coefficient of reliability. Thus, the main criterion becomes ease of administration and the degree of discrimination desired. The first system, a pass-fail dichotomy, is most advantageous when a rough measure is desired in a pass or fail situation. The fourth system, granting equal credit for each correct response at each level, gives the finest discrimination and is easier to use than either of the other two methods. The relative saving of time of pass-fail scoring in contrast to the fourth system is insignificant where there is an advantage in having as complete a record of this function as possible.

Compared to other tests in ease of administration, maintenance, variables demanding attention, trained personnel and initial cost, this test ranks among the first. Where large testing and selection programs are in operation with testing units scattered over a large area and working under varied conditions, this test leaves little to be desired.

****

The testing program of this study was carried on in conjunction with a project determining the reliability of various acuity test targets. (5) The author wishes to acknowledge the contribution of Lt.(jg) Backstrom, H(S), USNR, who cooperated in the administration of this test in the course of the other study.

REFERENCES

1. Zeethout, W. D., "Physiological Optics", the Professional Press, Inc., Chicago, Illinois, 1939.

2. Verhoeff, F. H., "Simple Quantitative Test for Acuity and Reliability of Binocular Stereopsis." Arch. Opth. 28:1000-1014, December 1942.

3. Bielschowsky, A., "Functional Disturbances of the Eyes." Arch. Opth., Vol. 15, No. 4, April 1936.

4. Betts, E. A., "Data on Visual Sensation and Perception Tests, Part III, Stereopsis." Keystone View Co., Meadville, Pa., 1940.

5. "A Comparison of the Reliability and Validity of Visual Acuity Test Targets." BuMed Project No. X-676(Av-357-p) by Lt. Trumbull, H(S), USNR, and Lt.(jg) Backstrom, H(S), USNR.

6. "Comparison of Ortho-Rater with Clinical Opthalmic Examinations." Final report on BuMed Project No. X-499(Av-263-p) by Comdr. Welpaw, (MC), USNR, and Lt. Imus, H(S), USNR.